

Dynamic ligand-based pharmacophore modeling and virtual screening to identify mycobacterial cyclopropane synthase inhibitors

CHINMAYEE CHOUDHURY^{a,b}, U DEVA PRIYAKUMAR^{b,*} and G NARAHARI SASTRY^{a,*}

^aCentre for Molecular Modelling, Indian Institute of Chemical Technology, Hyderabad, India

^bCentre for Computational Natural Sciences and Bioinformatics, International Institute of Information and Technology, Hyderabad, India

e-mail: deva@iiit.ac.in; gnsastry@gmail.com

MS received 5 November 2015; revised 11 February 2016; accepted 2 March 2016

Abstract. Multidrug resistance in *Mycobacterium tuberculosis* (*M. Tb*) and its coexistence with HIV are the biggest therapeutic challenges in anti-*M. Tb* drug discovery. The current study reports a Virtual Screening (VS) strategy to identify potential inhibitors of Mycobacterial cyclopropane synthase (CmaA1), an important *M. Tb* target considering the above challenges. Five ligand-based pharmacophore models were generated from 40 different conformations of the cofactors of CmaA1 taken from molecular dynamics (MD) simulations trajectories of CmaA1. The screening abilities of these models were validated by screening 23 inhibitors and 1398 non-inhibitors of CmaA1. A VS protocol was designed with four levels of screening *i.e.*, ligand-based pharmacophore screening, structure-based pharmacophore screening, docking and absorption, distribution, metabolism, excretion and the toxicity (ADMET) filters. In an attempt towards repurposing the existing drugs to inhibit CmaA1, 6,429 drugs reported in DrugBank were considered for screening. To find compounds that inhibit multiple targets of *M. Tb* as well as HIV, we also chose 701 and 11,109 compounds showing activity below 1 μ M range on *M. Tb* and HIV cell lines, respectively, collected from ChEMBL database. Thus, a total of 18,239 compounds were screened against CmaA1, and 12 compounds were identified as potential hits for CmaA1 at the end of the fourth step. Detailed analysis of the structures revealed these compounds to interact with key active site residues of CmaA1.

Keywords. Virtual screening; pharmacophore model; docking; tuberculosis; HIV; ADMET filters; drug repositioning; poly-pharmacology.

1. Introduction

Recent studies suggest that over two billion people, which is approximately one-third of the world population are infected with *M. Tb*^{1,2} highlighting the enormous healthcare challenges particularly for the developing countries. According to the ‘global tuberculosis report 2014’ by the World Health Organization (WHO), there were about 9.0 million incidences of tuberculosis (TB) in 2013.³ It is important to note that out of 1.5 million fatalities caused by this deadly disease, 0.36 million people were HIV positive.³ Treatment of TB for HIV infected patients is very complicated due to the incompatibilities of the currently available HIV and TB drugs.⁴ Hence, molecules that are able to inhibit multiple *M. Tb* targets as well as simultaneously act on both *M. Tb* and HIV are required to address the therapeutic challenges like drug resistance and coexistence with HIV. New targets and drug candidates are being explored by researchers across the globe to fulfill the

urgent need of new drugs for TB.^{5–7} These emerging targets include GlgE, enzymes involving mycolic acid cyclopropanation, DprE1/DprE2, MshC, HisG, AtpE, Def and methionine amino peptidase, etc., which are crucial for the actively growing mycobacteria. Few new chemical entities such as dioctylamine, benzothiazinone, dinitrobenzamides, dequalinium chloride, nitrobenzothiazole, diarylquinoline, TMC207, LBK-611 and 2,3-dichloro-1,4-naphthoquinones are in clinical trials.⁸ However, the new drug discovery strategies should include multidisciplinary approaches to identify and understand the structure and dynamic behavior of the potential drug targets.

Mycobacterial cyclopropane synthase (CmaA1) is one of the potential anti-TB drug targets, which is responsible for the maturation of mycolic acids by cyclopropanation, thus contributing towards the persistence and virulence of *M. Tb*. Our previous studies reported the conformational changes in the active sites of CmaA1 during the cyclopropanation reactions by performing molecular dynamics (MD) simulations on five model systems that represent different catalytic

*For correspondence

stages of the enzyme.⁹ The diverse conformational space of the active site residues of CmaA1 at various stages of the cyclopropanation using MD simulations, observed from this study prompted us to use this flexibility of CmaA1 active site in drug design. We generated dynamic structure-based pharmacophore models from the snapshots obtained from the MD trajectories of CmaA1 and compared their abilities to screen active inhibitors with that of the crystal structure-based model.¹⁰ It has been demonstrated that the pharmacophore models generated from the snapshots of the MD trajectory are much more efficient than the one generated from the single crystal structure. In this study, we present an effective strategy based on the development and validation of dynamic ligand-based pharmacophore models followed by virtual screening.

Considering the immediate need of new therapeutic agents to treat tuberculosis, we are interested to screen the existing drugs against CmaA1 inspired by the concept of 'Drug repurposing/Drug repositioning', wherein we look for new indications for the existing drugs. This is a less expensive and less time consuming process as well, due to the availability of the pharmacokinetic, toxicology and safety data. For screening purposes, the DrugBank¹¹ is taken as the first dataset in this study, especially considering the fact that the drug repurposing is emerging as one of the most promising strategies in drug discovery.^{12,13} (ChEMBL-MTb) which consists of all compounds, experimentally proven as inhibitors of any one *M. Tb* target reported in the ChEMBL database,¹⁴ is considered as the second dataset to be screened against CmaA1. The compounds that can inhibit multiple targets can more effectively combat drug resistance by mutation.¹⁵ A third dataset (ChEMBL-HIV) comprising all compounds, which are highly active against HIV was considered for screening to identify compounds that are able to bind and inhibit targets of both *M. Tb* and HIV. Therefore, these three datasets were considered in the study based on their drug-like nature, or activity against *M. Tb* or HIV.

Along with effective strategies for anti-TB drug discovery considering the resistance and coincidence with HIV, reliabilities of the techniques used during the process play a crucial role. In recent years, technological advances in computational methodologies like structure and ligand-based pharmacophore screening,^{16–19} docking,²⁰ 2D and 3D quantitative structure activity relationship (QSAR) modeling,²¹ etc., have offered medicinal chemists fast and cost-effective alternatives to the traditional high throughput screening to screen drug libraries against therapeutic targets. Consideration of the receptor flexibility using MD simulations in the VS methods and MD-based descriptors such as

MM-GBSA/PBSA energies as molecular descriptors add great value to the accuracy of inhibitor prediction.^{22,23} At the same time, all these methodologies have their own limitations as they are based on many approximations or demand very high computational power. For example, in QSAR modeling, we may generate large number of models to predict the activities, but it is very challenging to choose the right types of descriptors.²⁴ Most of the docking methods calculate interaction energies and free energies of binding by empirical methods, and they do not accurately account for the receptor flexibilities and employing MD simulations, while screening millions of compounds are computationally expensive. Since each individual method of VS has its own limitations, it is difficult to attain accuracy in the VS protocols unless we employ different screening filters like pharmacophore mapping, shape-based, ligand-based, chemical diversity, similarity based strategies. Incorporation of the target specific filters along with the general VS approaches is also very crucial.^{25–29}

In the current study, a systematic approach was designed employing dynamic ligand and structure-based pharmacophore screening followed by docking to screen the compound datasets (described above) against CmaA1. The pharmacokinetic properties like ADMET profiles of the molecules are some of the very important aspects to be considered while screening compounds to avoid dismissal of the drugs in the later stages of drug discovery.^{30,31} In silico prediction of these properties can rapidly analyze a set of molecules prior to synthesis and help prioritizing the molecules that can then be further investigated experimentally.^{32,33} So, ADMET filters were considered as the last levels of compound screening.

2. Computational

2.1 Generation and validation of ligand-based pharmacophore models

Our previous studies reported MD simulations on the following five model systems of CmaA1.^{9,10} **E-SAM**: Holo form of CmaA1 with SAM in the cofactor binding site (CBS). **E-SAM-S**: Holo form with SAM in the CBS and the substrate in the acetyl substrate binding site (ASBS). **E-SAHC-P**: Holo form with SAHC in the CBS and the cyclopropanated product in the ASBS. **E-SAHC**: Holo form with SAHC in the CBS. **E-SAHC-D**: Holo form with SAHC in the CBS and a CmaA1 inhibitor DDDMAB in the ASBS. These model systems represent various catalytic states of CmaA1.⁹ The conformational heterogeneity of the cofactor as obtained from these MD simulations was used to design dynamic

ligand-based pharmacophores. Eight snapshots were collected at every 5 ns interval from each of the five MD trajectories of the five model systems of CmaA1 (total 40 snapshots). The cofactors were extracted from each of these snapshots and average structures were generated for the extracted cofactors from each trajectory using “General tools” module of Schrodinger Suite.³⁴ All the cofactors extracted from one trajectory at every 5 ns were aligned to each other based on the adenine and sugar parts. The root mean square deviations (RMSD) among the snapshots extracted from a single trajectory were found to be $<2.5 \text{ \AA}$. Since the structures were quite similar to each other, average structures were calculated using uniform weighting method. Thus, we obtained five average structures of the cofactors from the five model systems. These five average structures of the cofactors along with the one extracted from the static crystal structure (1KPH) were used to generate the ligand-based pharmacophore models using the Phase module of the Schrodinger program.³⁵ Each model consisted of six types of chemical features. The number of features present in each model varied from 8 to 11. A set of 23 CmaA1 inhibitors with reported MIC values ranging from 0.0125-12.5 $\mu\text{g/mL}$ ³⁶ were used for verifying the abilities of the ligand-based pharmacophore models to screen active CmaA1 inhibitors (scheme S1). All these compounds were energy minimized using the default parameters of LigPrep module of Schrodinger Suite.³⁷ The default parameters used for ligand preparation included the OPLS_2005 force field and to retain the existing charge. A maximum of 10 conformations per molecule were generated during the matching. The matching criteria were assigned as the compounds must match at least 4 features of a model. A set of 1398 *M. Tb* inactive compounds reported in ChEMBL database (used in the previous study)¹⁰ were found to be within the molecular weight range of 180-400 and consisting 12 to 27 heavy atoms (similar to that of the 23 active compounds, SAM and SAHC). These 1398 compounds were then screened against all e-Pharmacophore models using the same criteria to check if these models screen any inactive compounds.

2.2 Virtual Screening

2.2a Preparation of dataset: Three sets of compounds were used for VS against CmaA1. First one consisted of all the 6,429 drugs (approved/illicit/withdrawn/nutraceuticals) reported in DrugBank. 701 unique compounds showing activity below 1 μM range on *M. Tb* cell lines reported in ChEMBL database constituted the second set, and 11,109 compounds showing activity below 1 μM range on various HIV cell lines were

collected from ChEMBL database as the third set. All these 18,239 compounds were energy minimized using the default parameters of LigPrep module of Schrodinger Suite. The individual datasets mentioned above are referred as ‘DrugBank’, ‘ChEMBL-Mtb’ and ‘ChEMBL-HIV’ respectively. Detailed information on the procedure and date of downloading the compounds and preparation of the above datasets have been given in as List S1 of the supplementary information.

2.2b Screening: The three datasets were screened individually to identify potential CmaA1 inhibitors from each set. Four different levels of filters were used for screening each of these datasets. The ligand-based pharmacophore models were chosen as the first level filters for VS as they have more number of features and hence can screen diverse compounds. The ‘Advanced Pharmacophore Screening’ protocol of Phase module of Schrodinger suite was used to map the active compounds with the pharmacophore models. Five conformations were generated for each rotatable bond in the active compounds and upto 100 conformations per compound were generated. Compounds those match to a minimum of six pharmacophoric features of a ligand-based model were allowed to get screened in. The compounds screened by all the five ligand-based pharmacophore models were subjected to the second level filter. Five structure-based pharmacophore models that have been generated and validated in our earlier study were used as the second level of screening as they can screen target specific compounds.¹⁰ All the screening criteria were same as the first filter, but here thorough conformational sampling was used and the minimum number of pharmacophore sites the molecules must match to be screened in was set to be 4. All the compounds that could map to any four features of each of the five structure based pharmacophore models were subjected to the next level of screening *i.e.*, docking. The selected compounds were docked to the active sites of the corresponding protein structures of the five structure-based pharmacophore models. Compounds having reactive functional groups were opted out before the Glide docking.³⁸ Glide energy grids were generated for each snapshot to define the active site as a cubic box of $12*12*12 \text{ \AA}^3$ around the cofactors. Docking was performed in two sub-steps *i.e.*, the simple precision (SP) docking and an extra precision (XP) docking.³⁹ Top 50% compound poses ranked according to Glide SP score were subjected to Glide XP docking and top 50% compounds ranked according to the XP scores were retained as top hits. The common compounds in the top 25% of all the five docking results were selected for the ADMET/Drug likeness property calculation with the

(H-bond) acceptor (A), H-bond donor (D), hydrophobic sites (H), negative ionic sites (N), positive ionic sites (P) and aromatic rings (R). D features were represented as projected points, located at the corresponding A positions in the binding site. Projected points allow the possibility for structurally dissimilar active compounds to form H-bonds to the same location, regardless of their point of origin and directionality. All the ligand-based models basically consisted of similar features except an extra P feature for SAM and an extra H feature for SAHC. But the inter-feature distances, spacing and orientations largely vary depending on the conformation of these cofactors in the binding pocket. The two R features of the adenine part represent the potential of making π - π interaction with the side chain of F142 residue of CmaA1 as observed in our previous studies.⁹ The adenine part also consists of two A features which mostly interact with E124 and Y16 of CmaA1 and one D feature that interacts with W123, the sugar part

contains two D and two A features situated on the two -OH groups in all the models which have potential to make H-bonds with the residues G72, G74, L95 and Q99. The methionine part has a P feature for the ligand-based models of E-SAM and E-SAM-S while the corresponding homocysteine parts in E-SAHC, E-SAHC-D and E-SAHC-P have H features. The terminal parts of all the ligand-based models contain the P and N features which interact with I136, Y33, S34 and T94. The abilities of the generated models were verified by screening a set of 23 reference compounds showing CmaA1 inhibitory activities in $\mu\text{g/mL}$ range as in the previous study.¹⁰ Figure 2 shows the five ligand-based models with the most active reference compound C1 mapped to them.

The active compounds considered for validation of the models are the anti-tubercular drug thiacetazone and its clinical analogues that are shown to cause inhibition of cyclopropanation in various mycobacterial strains

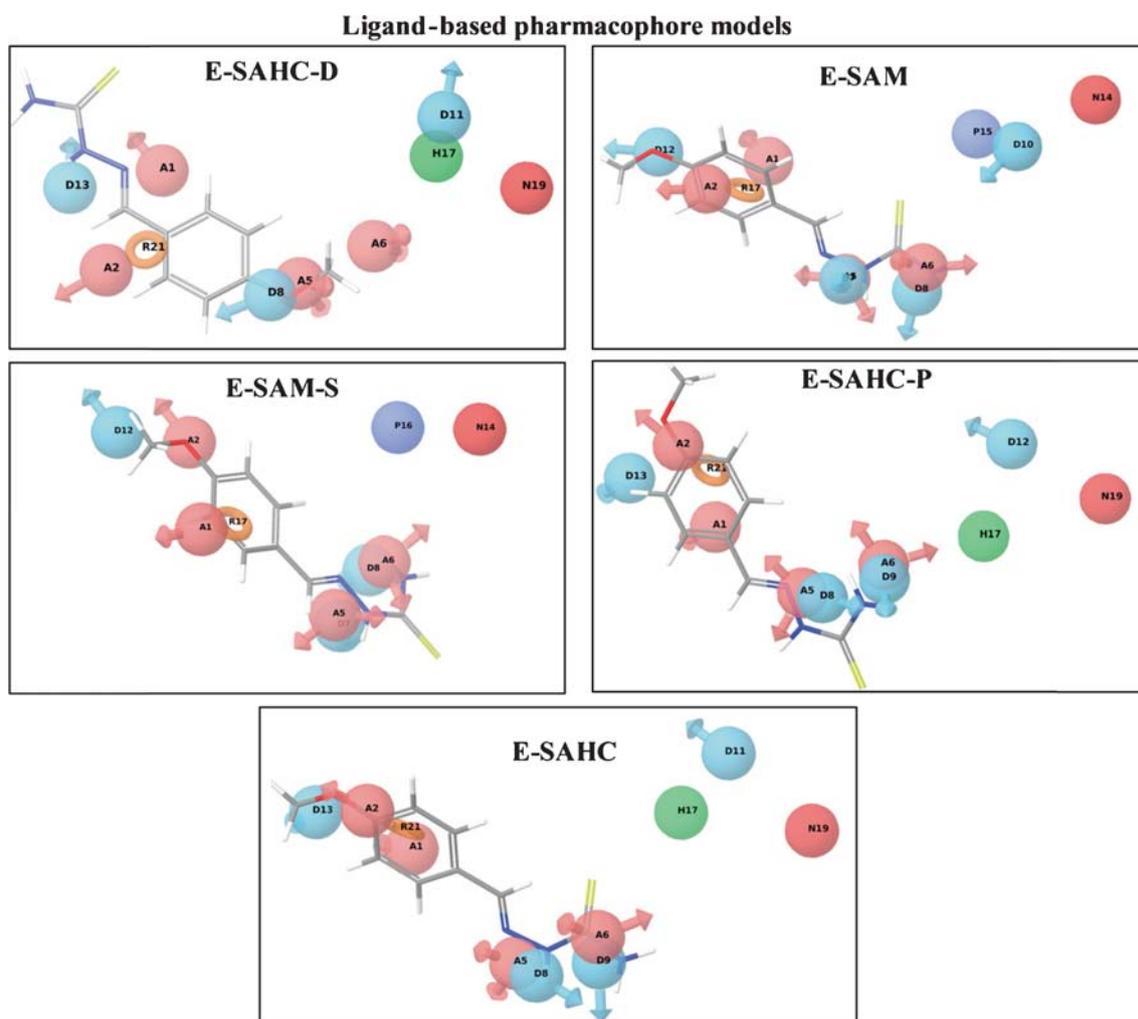


Figure 2. The most active reference compound C1 mapped with all the selected ligand-based pharmacophore models. Colour codes for the pharmacophoric features are as follows. Cyan: D, Pink: A, Red: N, Blue: P, Green: H and Orange: R. Same colour code for the features is followed for all the other figures.

Table 1. Comparison of number of inhibitors and non-inhibitors screened by the ligand-based pharmacophore models and ranges of their fitness scores.

Model	% of inhibitors screened	Fitness Score Range	% of non-inhibitors screened	Fitness Score Range
E-SAHC-D	65	0.68–1.15	10.2	0.25–1.04
E-SAM	96	1.65–1.90	8.1	0.26–1.29
E-SAM-S	83	0.57–1.38	0.14	0.06–1.09
E-SAHC-P	91	0.62–1.45	10.8	0.24–1.06
E-SAHC	91	0.46–1.69	12.4	0.20–1.05

by directly binding cyclopropane synthase enzymes.³⁵ The ligand-based models were able to screen upto 22 out of the 23 active compounds. Since the number of features present in these models were more (8-11), they could screen more compounds when the matching criteria was kept to minimum four features for a molecule. The ligand-based model generated from the co-crystallized SAHC from the static crystal structure was found to have features similar to those of the dynamics-based models. But, it could screen only four active compounds. The ligand-based models and the reference compounds screened by them have been shown in table S1. When the 1398 non-inhibitors of *M. Tb* were screened using the five ligand-based models, it was found that the models screen very less percentage of non-inhibitors as compared to the inhibitors (table 1).

Table 1 also shows that the fitness scores for the non-inhibitors are lower than those of the inhibitors, demonstrating the ability of the models to discriminate the inhibitors from the non-inhibitors. Hence, the five ligand-based pharmacophore models generated from the MD trajectories were used as the first level filters in the VS process.

3.2 Virtual Screening

3.2a Design of the VS protocol: The literature report several methods to screen potential compounds for potential targets. Although the pharmacophore screening is one of the fast and cost effective techniques, it has its own limitations which have been discussed elsewhere.⁴¹ Similarly, other screening techniques based on docking use severe approximations to estimate the binding affinity of the ligands for a particular protein. Also, docking-based methods do not consider receptor flexibility to save time and computational cost. Hence, it would be wise to design the VS protocols employing several techniques so that all their strengths can be exploited, as well as their limitations will be complemented by each other. So, here we employed ligand and structure-based pharmacophore models incorporating dynamics of the receptor, docking with

multiple receptor conformation and ADMET filters for screening the dataset of our interest. Figure 3 shows the step by step VS process used in our study. Each dataset was screened individually as they were considered to address three different purposes. No common compounds were found among the datasets.

3.2b First level filter: Dynamic ligand-based pharmacophore screening: The first level filter used for our study was the five ligand-based pharmacophore models. These models are based on the diverse conformations of the cofactors in the binding site of CmaA1 at various stages of the enzyme catalysis. They have 8-11 features and hence can screen structurally diverse compounds satisfying the ligand-based requirements of a CmaA1 binder. The pharmacophoric features represent the key non-covalent interactions of the cofactors with the active site residues of CmaA1 at different catalytic stages of cyclopropanation. Due to high conformational difference of the cofactors at different stages, their pattern of interaction with the active site residues is also altered. For example, in the E-SAM stage, SAM binding is stabilized by strong H-bonds with Ile136, Gly137, Glu124 and Gly72. But in presence of the substrate, in the E-SAM-S stage, SAM changes its conformation so that the CH₃ group attached to the S⁺ of SAM tends to come closer to the C=C of the substrate and due to this conformational change the previous H-bonds are lost and new H-bond are formed with ASP70 and Thr78.⁹ Leu95 which was making H-bonds with N3 and O2' of SAM in the E-SAM system makes H-bond with N7 of SAM in the E-SAM-S system because of its altered orientation.⁹ We feel it is very essential to consider these diverse interaction patterns of the cofactors while screening compounds for CmaA1. The three datasets were screened individually by the five dynamic ligand-based models and the common set of compounds obtained by all the five models were selected for the next level screening. The compounds probability of binding of these screened compounds are expected to be higher as they are able to attain conformations so that the orientations of their

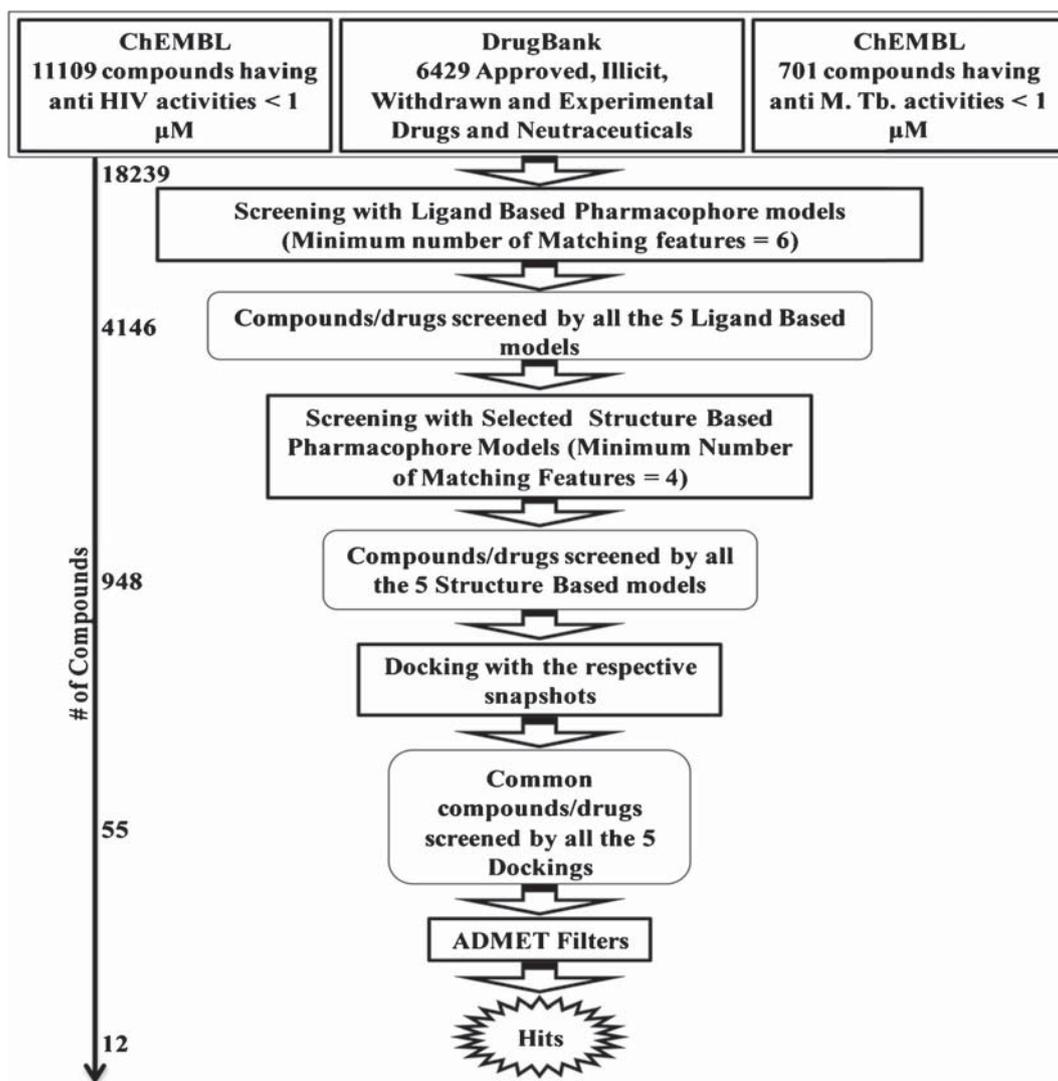


Figure 3. Schematic representation of the step by step virtual screening process.

pharmacophoric features match the multiple interaction pattern of SAM/SAHC at different stages of cyclopropanation. The first level screening returned 1438, 166 and 2542 compounds from the datasets DrugBank, ChEMBL-MTb and ChEMBL-HIV respectively, thus a total of 4146 compounds.

3.2c Second level filter: Dynamic structure-based pharmacophore screening: The next level filter applied was the five structure-based models selected in our previous study.¹⁰ These models are based on the interaction of the cofactors with the active site residues of CmaA1, hence are more specific for the target. In the selected structure-based models, the D features were formed near the residues G72, T94, Q99 which make H-bonds with the sugar moiety of SAM/SAHC and also with E124, Y16, which make H-bonds with the adenine moiety of SAM/SAHC. In most cases the A features

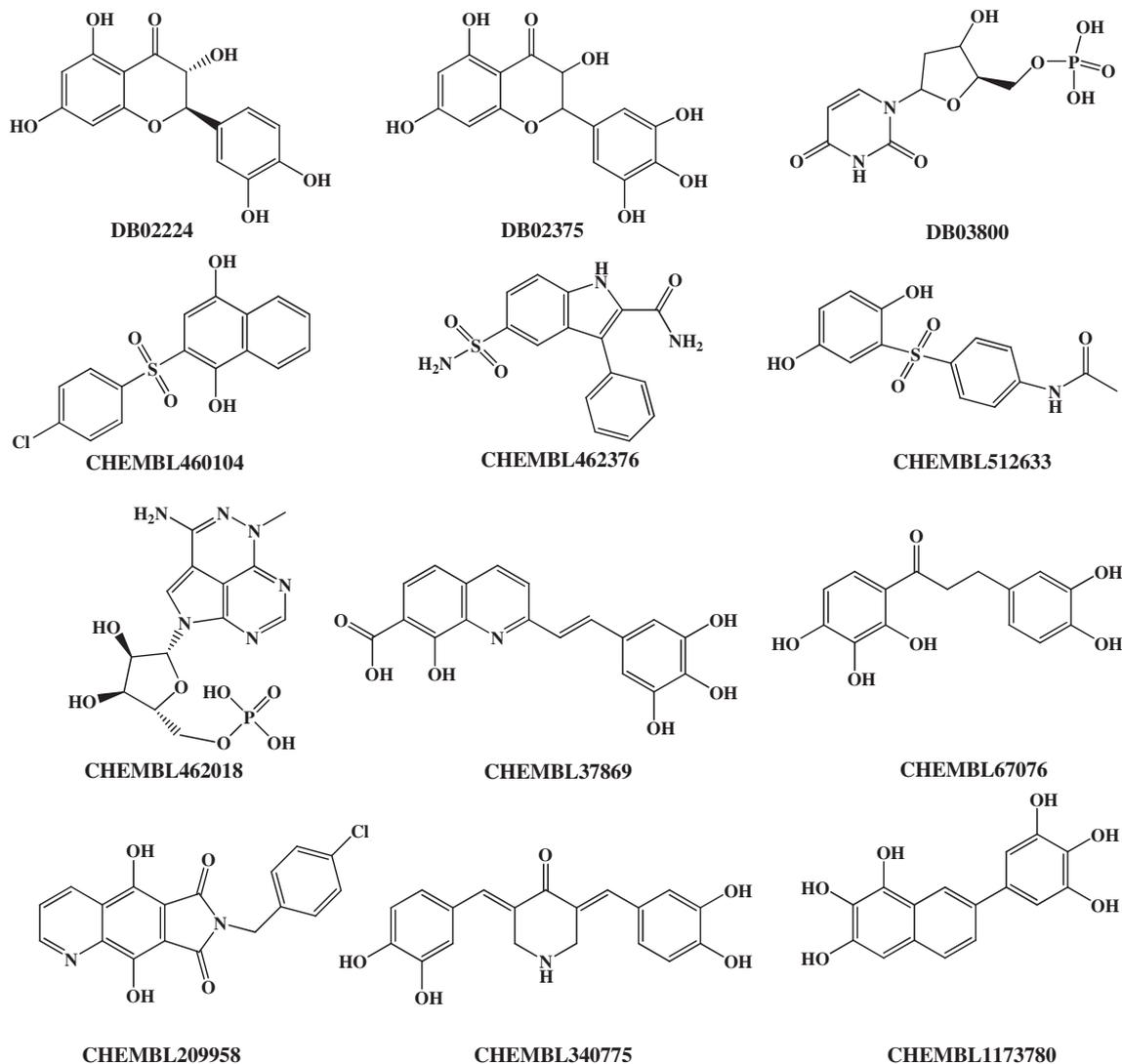
were found near L95, Q99, G74 which make H-bonds with the sugar moiety of the cofactors and the residues W123 which makes H-bonds to the adenine part of the cofactors. The P and N features are found mostly near the residues I136 and Y33, S34 respectively which show electrostatic interaction with the polar terminal part of the cofactors. A thorough conformational sampling was used at this stage of screening and a unique set of 948 compounds in total and 532, 33 and 383 compounds from the datasets DrugBank, ChEMBL-MTb and ChEMBL-HIV respectively were obtained at the second level.

3.3 Third level filter: Docking

The compounds screened in the previous level were subjected to the third level screening, *i.e.*, docking. All the 948 compounds were docked to the active sites

of the parent snapshots of the selected structure based models *i.e.*, the snapshots from the MD trajectories of the model systems E-SAHC-D at 15 ns and 20 ns, E-SAM at 30 ns and E-SAHC at 10 ns and 35 ns. These snapshots were shown to screen maximum numbers of CmaA1 active compounds.⁶ Top scoring compounds of each docking run with each snapshot were analyzed. The compounds present in the top 25% of all the five snapshots were passed to the next level of screening. Thus the screened compounds are expected to bind to many optimal conformations of the flexible active site of CmaA1. A total of 55 compounds *viz.*, 30 from DrugBank, 8 from ChEMBL-MTb and 17 from ChEMBL-HIV were screened as the top scoring hits screened by all the five snapshots. Table S2 shows the docking scores of the 55 compounds screened at the third level with all the five selected snapshots of CmaA1.

3.3a Fourth level filter: ADMET properties: The screened compounds were then subjected to calculation of ADMET properties with QuickProp module of Schrodinger which predicts many significant and pharmacologically relevant properties to estimate the drug likeliness of a given molecules. One can compare certain properties of a particular molecule with the given ranges of those of 95% of known drugs. Also, QuickProp can identify the presence of 30 types of reactive functional groups that may cause false positives in VSs. The important properties that are calculated and can be compared with the ranges of known drugs are MW, dipole, IP, EA, SASA, FOSA, FISA, PISA, WPSA, PSA, volume, #rotor, donorHB, acptHB, glob, QPpolrz, QPlogPC16, QPlogPoct, QPlogPw, QPlogPo/w, logS, QPlogKhsa, QPlogBB, #metabol, etc. The descriptions of all these descriptors are listed in List S2. We have prioritized our screened compounds



Scheme 1. Compounds selected structures of all the selected compounds from all the three datasets taken.

based on the number of property or descriptor values that fall outside the 95% range of similar values for 95% of known drugs (#stars) calculated by QuickProp, Hence a smaller #stars suggests that a molecule is more

Table 2. Pharmacophore fitness and docking scores of the 12 screened hits with the ligand and structure based pharmacophore models and the snapshots.

Pharmacophore fitness scores of the hit compounds with all the 5 ligand based pharmacophore models						
Hit Compounds	Dataset	E-SAHC-D	E-SAM	E-SAM-S	E-SAHC-P	E-SAHC
CHEMBL460104	ChEMBL-Mtb	1.16	1.53	1.30	1.32	1.34
CHEMBL462376		1.20	1.50	1.15	1.28	1.40
CHEMBL512633		1.42	1.53	1.19	1.33	1.10
DB02224	DrugBank	1.34	1.2	1.29	1.28	1.27
DB02375		1.29	1.16	1.29	1.39	1.28
DB03800		1.42	1.67	1.49	1.45	1.37
CHEMBL37869	ChEMBL-HIV	1.03	1.00	2.15	1.32	1.29
CHEMBL67076		0.62	1.02	1.36	1.71	1.54
CHEMBL209958		0.71	1.26	1.38	1.4	1.36
CHEMBL340775		1.17	1.31	1.41	1.34	1.61
CHEMBL1173780		0.62	0.69	1.33	1.19	1.63
CHEMBL462018		0.89	1.15	0.82	1.29	1.24
C1	Reference* ³¹	0.73	1.87	0.96	0.79	1.22
Pharmacophore fitness scores of the hit compounds with all the 5 structure based models						
Hit Compounds	Dataset	E-SAHC-D (15 ns)	E-SAHC-D (20 ns)	E-SAM (30 ns)	E-SAHC (10 ns)	E-SAHC (35 ns)
CHEMBL460104	ChEMBL-Mtb	1.66	1.49	1.43	1.36	1.47
CHEMBL462376		1.58	1.56	1.32	1.28	1.17
CHEMBL512633		1.59	1.58	1.24	1.40	1.45
DB02224	DrugBank	1.42	1.48	1.73	1.54	1.37
DB02375		1.70	1.59	1.52	1.45	1.61
DB03800		1.57	1.37	1.26	1.50	1.57
CHEMBL37869	ChEMBL-HIV	1.67	1.47	1.51	1.38	1.40
CHEMBL67076		1.27	1.41	1.68	1.67	1.52
CHEMBL209958		1.35	1.31	1.21	1.56	1.31
CHEMBL340775		1.44	1.38	1.50	1.60	1.43
CHEMBL1173780		1.69	1.55	1.35	1.71	1.58
CHEMBL462018		1.37	1.25	1.34	1.65	1.42
C1	Reference* ³¹	1.25	0.76	1.25	0.97	1.06
XP docking scores of the hit compounds with the 5 snapshots corresponding to the structure based models						
Hit Compounds	Dataset	E-SAHC-D (15 ns)	E-SAHC-D (20 ns)	E-SAM (30 ns)	E-SAHC (10 ns)	E-SAHC (35 ns)
CHEMBL460104	ChEMBL-Mtb	-9.11	-9.31	-10.39	-10.34	-11.83
CHEMBL462376		-8.00	-10.45	-8.39	-14.46	-11.71
CHEMBL512633		-7.20	-9.85	-7.72	-11.47	-10.87
DB02224	DrugBank	-8.42	-8.69	-9.53	-11.26	-10.83
DB02375		-10.00	-9.68	-8.91	-11.56	-12.23
DB03800		-13.02	-9.83	-8.93	-12.08	-9.28
CHEMBL37869	ChEMBL-HIV	-9.76	-6.70	-10.65	-12.09	-13.99
CHEMBL67076		-9.17	-8.46	-9.01	-12.15	-11.78
CHEMBL209958		-7.06	-6.72	-7.90	-8.89	-9.38
CHEMBL340775		-9.52	-8.36	-8.20	-10.19	-11.29
CHEMBL1173780		-8.68	-10.39	-9.75	-10.40	-13.19
CHEMBL462018		-8.02	-8.03	-8.26	-13.13	-12.32
C1	Reference* ³¹	-3.99	-4.54	-5.68	-5.29	-6.99

*Fitness and docking scores of the most active compound C1 used for validation of the models.³¹

drug-like than molecules with more #stars. We screened all the compounds that have passed our previous filters which have #star as 0. Thus we obtained a total of 12 compounds that could pass this filter, three DrugBank compounds, three ChEMBL-MTb compounds and six ChEMBL-HIV compounds. Tables S3–S5 list all the QuickProp properties of the selected compounds screened at the fourth level.

The fitness scores of the screened compounds with the five ligand-based models and five structure-based models have been given in table 2. Table 2 also gives the XP docking scores of the 12 compounds with

the CmaA1 snapshots corresponding to the structure based pharmacophore models. The fitness scores of the screened hits with the ligand-based pharmacophore models are mostly higher than the ranges of fitness scores of the inactive compounds (mentioned in table 1) for a particular model. A quick look at table 2 also reveals that the fitness scores of the screened compounds are higher than those of the most active compound C1 considered for validation of the models in this study.³⁵ Similar trends were also found for the fitness scores of the screened hits with the structure based pharmacophore models. The upper ranges of the fitness scores of the inactive compounds for the structure based

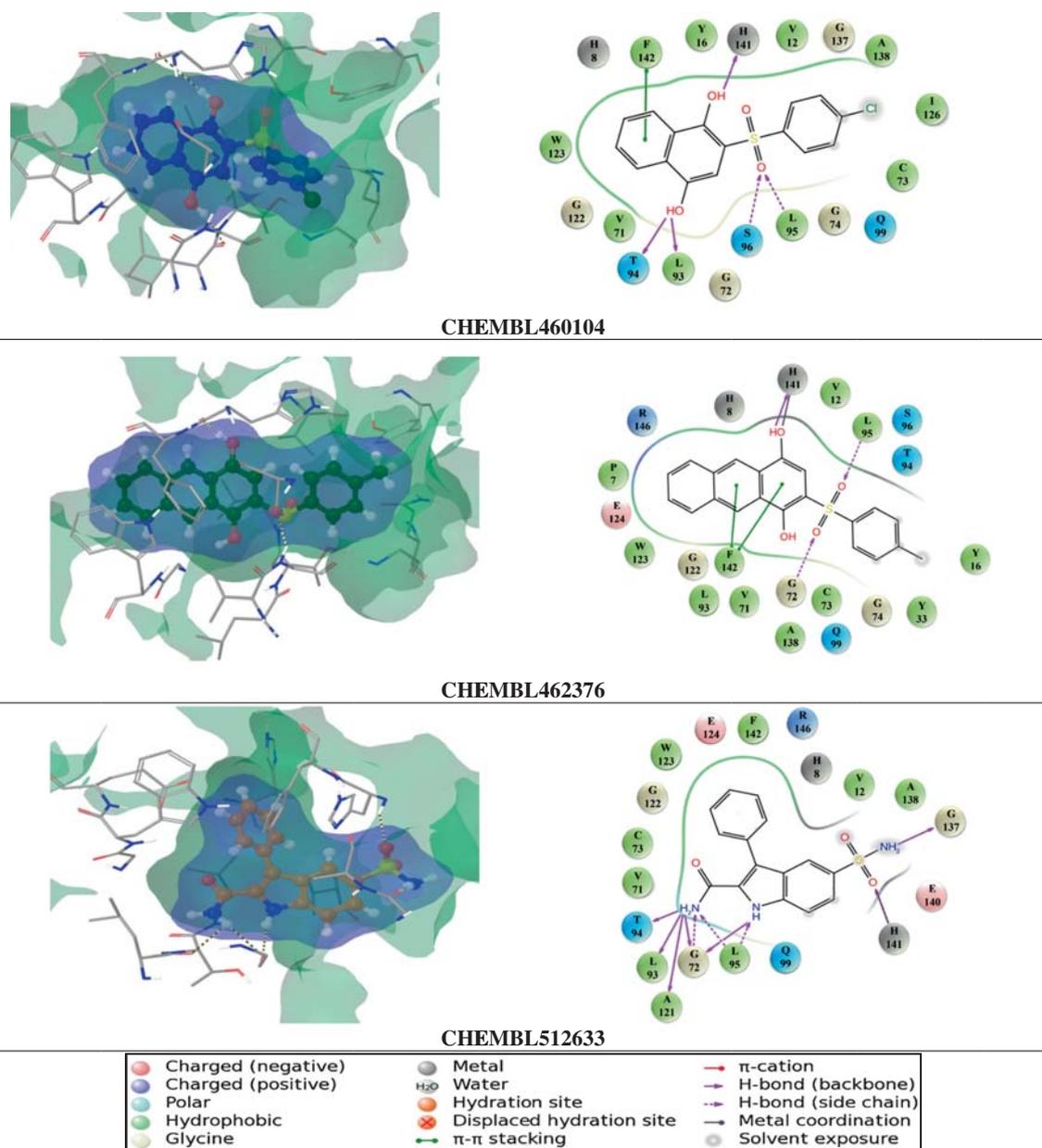


Figure 4. Interaction of the screened DrugBank compounds with the CmaA1. Among the five docking poses with 5 selected snapshots, the complexes with highest docking scores have been shown.

very low IC_{50} values of 3.69 and 31.4 nM against *M. Tb* FabH. Our docking studies reveal that the aromatic rings of both these compounds make π - π interactions with the F142 residue and make H-bonds with the residues T94 and L95. CHEMBL460104 makes an additional π - π interaction with the H141 residue. The third ChEMBL-MTb compound CHEMBL512633 screened in our study was synthesized and biologically evaluated by Guzel *et al.* This compound showed a K_i of 7.2 and 7.5 nM against *M. Tb* recombinant carbonic anhydrases Rv3273 and Rv1284, respectively, and showed selectivity for the *M. Tb* targets than their human homologues.⁴⁵ This compound makes H-bonds with the E124 and H141 residues and the aromatic system makes π - π interactions with the F142 residue. The study identified these compounds as potential *M. Tb* CmaA1 inhibitors, as well as explored their new applications. These three compounds are expected to inhibit CmaA1 along with the other *M. Tb* targets.

Three DrugBank compounds (figure 5) were identified among the top 12 hits. These can be considered for further experimental assays on CmaA1 for drug repositioning. One of the three selected DrugBank compounds is (2s, 3s)-trans-dihydroquercetin (DB02224)

which belongs to the class flavonoids with no information available about the target in DrugBank. This compound has high similarity with the approved drug Hesperetin (0.907 Tanimotto coefficient) which is a sterol o-acyltransferase 1 inhibitor used for lowering cholesterol. This compound binds to *M. Tb* CmaA1 by making-bonds with the residues W123, D124, H141 and Q99 residues. The electron rich aromatic ring also makes π - π interactions with the F142. The second screened DrugBank compound is Myricetin (DB02375) which also belongs to flavonoids class and act as inhibitors of multidrug resistance-associated protein1, which may be useful in managing of antimicrobial drug resistance.⁴⁶ This compound makes H-bonds with the residues Y16, G74, S96, Q99 and W123. It also makes π - π interactions with the F142 of CmaA1. The third DrugBank compound is an experimental small molecule 2'-deoxyuridylic acid (DB03800) belonging to the class carbohydrate conjugates which binds to proteins from a wide range of organisms including deoxyuridine 5'-triphosphate nucleotidohydrolase of *M. Tb* and thymidylate synthase of human. This compound binds to CmaA1 by making H-bond interactions with H8, G72, T94, L95, Q99, W123 and E124.

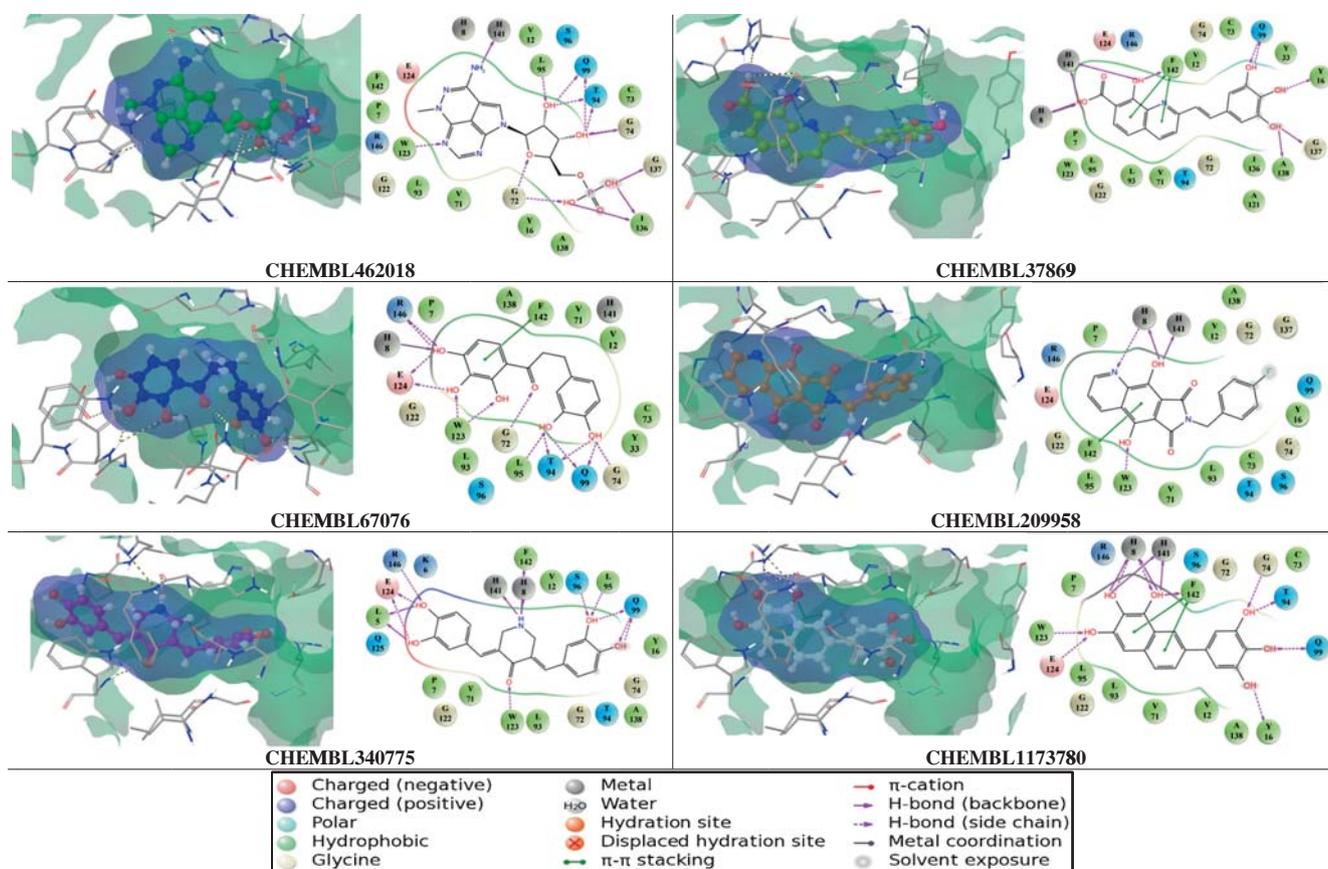


Figure 6. Interaction of the screened ChEMBL-HIV compounds with the CmaA1. Among the five docking poses with 5 selected snapshots, the complexes with highest docking scores have been shown.

Among the 6 screened ChEMBL-HIV compounds (figure 6), the first one is ChEMBL37869, which shows an IC_{50} of 300 nM against HIV integrase as reported by Zouhri *et al.*, who performed structure activity relationship studies and tested this compound for inhibition of HIV-1 integrase and replication of HIV-1 in cell culture. This compound makes H-bonds with the residues H8, Y33, Q99, G137 and H141 and makes π - π interactions with F142.

The compound ChEMBL67076 has been reported to inhibit the HIV-1 integrase with an IC_{50} of 1 μ M. This compound makes H-bonds with H8, G72, T94, L95, Q99 and E124 and π - π interactions with F142.¹⁶ The compound ChEMBL209958 was synthesized and the *in vitro* structure bioactivity study was performed by Fardi *et al.*⁴⁷ This compound shows an IC_{50} of 80 nM against HIV-1 integrase. Our docking study shows that this compound makes H-bonds with W123 and H141 and π - π interactions with F142 of CmaA1. The compound ChEMBL340775 was synthesized and tested for HIV-1 integrase inhibitory activity by Artico *et al.*, in 1998.⁴⁸ This compound showed an IC_{50} of 200 nM against the HIV-1 integrase. This binds to CmaA1 by making H-bonds with the residues L5, Q99, E124 and R146. The compound ChEMBL1173780 is a 13 hydroxylated 2-arylnaphthalene synthesized and tested for its inhibitory activities against HIV-1 integrase by Maurin *et al.*, in 2010.⁴⁹ This compound showed IC_{50} value of 500 nM against HIV-1 integrase. In our docking study it was found to bind to CmaA1 by making H-bond interactions with H8, Y33, G74, T94, Q99 and H141 and π - π interactions with F142. The compound ChEMBL462018 is one of the 6-N-acyltricyclic analogues synthesized and tested by Porcari *et al.*, showed IC_{50} of 40 nM against the reverse transcriptase activity in cells acutely infected with HIV-1.⁵⁰ This compound binds to CmaA1 by making H-bonds with the residues H8, Y33, G72 and E124 of CmaA1.

The compounds screened by the dynamics-based pharmacophore models and docking mostly interact with the residues H8, Y33, G72, G74, T94, L95, Q99, W123, E124 and H141 which were found to be important residues for the cofactor binding. Hence, these compounds are expected to be potential inhibitors of CmaA1 by competing with the natural cofactors for binding into the cofactor binding pocket of CmaA1.

4. Conclusions

Ligand-based pharmacophore models have been generated from the snapshots obtained from the MD simulation trajectories of five model systems of CmaA1

representing various stages of cyclopropanation taken from a previous study and the reported crystal structure of CmaA1. The performance of these pharmacophore models were validated by mapping 23 active and 1398 inactive reference compounds of CmaA1. The ligand-based pharmacophore models generated from the averaged structures of the cofactors extracted from the snapshots were able to screen upto 22 out of 23 CmaA1 active compounds and a very less percentage of inactive compounds. Performance of the dynamics-based models was found to be better than the model obtained from the conformation of SAHC in the crystal structure which could screen only four active compounds. A novel VS workflow was designed with four levels of filters *viz.*, ligand-based pharmacophore screening, structure based pharmacophore screening, docking and ADMET filters. A dataset containing a total of 18,239 compounds (6,429 drugs reported in DrugBank, 701 and 11,109 compounds showing activity below 1 μ M range on *M. Tb* HIV cell lines respectively collected from ChEMBL database) was screened using the VS workflow. The 12 screened compounds were found to bind to the CmaA1 active site by interacting with the residues H8, Y33, G72, G74, T94, L93, L95, Q99, W123, E124 and H141. These residues not only have important roles for the binding of the natural substrates of CmaA1, but were also found to undergo conformational changes that are necessary during the cyclopropanation reaction. Hence, the screened compounds maybe effective inhibitors of CmaA1 out of which 6 compounds may act as dual inhibitors of HIV and *M. Tb*. Experimental screening of these hits to verify their CmaA1 inhibitory activities are in progress.

Supplementary Information (SI)

Scheme S1 showing the reference CmaA1 inhibitors, List S1 gives the details of the dataset download and preparation, table S1 giving the details (number of compounds matched, docking and fitness scores, etc.) of all the ligand-based pharmacophore models, table S2–S5 giving the docking scores and QuickProp Descriptors of all screened compounds and List S2 giving the descriptions of the Quickprop descriptors are provided at www.ias.ac.in/chemsci.

Acknowledgements

CC thanks Department of Science and Technology (DST), New Delhi for financial assistance through INSPIRE fellowship. UDP thanks DAE-BRNS for financial assistance. GNS thanks CSIR, New Delhi for financial support in the form of XII five year project (GENESIS).

References

1. Balganesch T S, Alzari P M and Cole S T 2008 *Trends Pharmacol. Sci.* **29** 576
2. WHO 2014 Global Tuberculosis Report. <http://www.who.int/tb/en/> (accessed on 20/06/2015)
3. Zumla A, George A, Sharma V, Herbert N and Ilton B M 2013 *Lancet* **382** 1765
4. Varghese G M, Janardhanan J, Ralph R and Abraham O C 2013 *Curr. Infect. Dis. Rep.* **15** 77
5. Mdluli K, Kaneko T and Upton A 2015 *Cold Spring Harb. Perspect. Med.* doi: 10.1101/cshperspect.a021154
6. Kandel D D, Raychaudhury C and Pal D 2014 *J. Mol. Model.* **20** 2164
7. Raychaudhury C, Kandel D D and Pal D 2014 *Croat. Chem. Acta* **87** 39
8. Lamichhane G 2011 *Trends Mol. Med.* **17** 25
9. Choudhury C, Priyakumar U D and Sastry G N 2014 *J. Struct. Biol.* **187** 38
10. Choudhury C, Priyakumar U D and Sastry G N *J. Chem. Inf. Model.* **55** 848
11. Wishart D S, Knox C, Guo A C, Shrivastava S, Hassanali M, Stothard P, Chang Z and Woolsey J 2006 *Nucleic Acids Res.* **34** D668
12. Tobinick E L 2009 *Drug News Perspect.* **22** 119
13. Bohari M H and Sastry G N 2012 *J. Mol. Model.* **18** 4263
14. Bento A P, Gaulton A, Hersey A, Bellis L J, Chambers J, Davies M, Krüger F A, Light Y, Mak L, McGlinchey S, Nowotka M, Papadatos G, Santos R and Overington J P 2014 *Nucleic Acids Res.* **42** D1083
15. Kinnings S L, Liu N, Buchmeier N, Tonge P J, Xie L and Bourne PE 2009 *PLoS Comput. Biol.* **5** e1000423
16. Carlson H A, Masukawa K M, Rubins K, Bushman F D, Jorgensen W L, Lins R D, Briggs J M and McCammon J A 2000 *J. Med. Chem.* **43** 2100
17. Meagher K L and Carlson H A 2004 *J. Am. Chem. Soc.* **126** 13276
18. Meagher K L and Carlson H A 2005 *Proteins Struct. Funct. Bioinf.* **58** 119
19. Damm K L and Carlson H A 2007 *J. Am. Chem. Soc.* **129** 8225
20. Kitchen D B, Decornez H, Furr J R and Bajorath J 2004 *Nat. Rev. Drug Discov.* **3** 935
21. Kubinyi H 1997 *Drug Discov. Today* **2** 457
22. Srivastava H K and Sastry G N 2012 *J. Chem. Inf. Model.* **52** 3088
23. Carlson H A 2002 *Curr. Opin. Chem. Biol.* **6** 447
24. Srivastava H K, Choudhury C and Sastry G N 2012 *Med. Chem.* **8** 811
25. Badrinarayan P and Sastry G N 2011 *Comb. Chem. High Thr. Scr.* **14** 840
26. Badrinarayan P and Sastry G N 2012 *J. Mol. Graph. Modell.* **34** 89
27. Badrinarayan P and Sastry G N 2014 *PLoS One* **9** e113773
28. Reddy A S, Pati S P, Kumar P P, Pradeep H N and Sastry G N 2007 *Curr. Protein Pept. Sci.* **8** 329
29. Badrinarayan P and Sastry G N 2013 *Curr. Pharm. Des.* **19** 4714
30. Selick H E, Beresford A P and Tarbit M H 2002 *Drug Discov. Today* **7** 109
31. Kubinyi H 2003 *Nat. Rev. Drug Discov.* **2** 665
32. de Waterbeemd H V and Gifford E 2003 *Nat. Rev. Drug Discov.* **2** 192
33. Oprea T I, Davis A M, Teague S J and Leeson P D 2001 *J. Chem. Inf. Comput. Sci.* **41** 1308
34. Schrödinger Release 2015-4: Maestro, version 10.4, 2015, Schrödinger, LLC, New York, NY
35. Dixon S L, Smondyrev A M, Knoll E H, Rao S N, Shaw D E and Friesner R A 2006 *J. Comput. Aided. Mol. Des.* **20** 647
36. Anuradha A, Trivelli X, Guérardel Y, Dover L G, Besra G S, Sacchettini J C, Reynolds R C, Coxon G D and Kremer L 2007 *PLoS One* **12** e1343
37. LigPrep, version, 2.5 2012 Schrödinger, LLC, New York, NY
38. Glide, version, 5.8 2012, Schrödinger, LLC, New York, NY
39. Friesner R A, Murphy R B, Repasky M P, Frye L L, Greenwood J R, Halgren T A, Sanschagrin P C and Mainz D T 2006 *J. Med. Chem.* **49** 6177
40. QikProp version 3.5 2012, Schrödinger, LLC, New York, NY
41. Yang S Y 2010 *Drug Discov. Today* **15** 444
42. Saha S and Sastry G N 2015 *J. Phys. Chem. B* **119** 11121
43. Mahadevi A S and Sastry G N 2013 *Chem. Rev.* **113** 2100
44. Alhamadsheh M M, Waters N C, Sachdeva S, Lee P and Reynolds K A 2008 *Bioorg. Med. Chem. Lett.* **18** 6402
45. Güüzöl Ö, Maresca A, Scozzafava A, Salman A, Balaban A T and Supuran C T 2009 *J. Med. Chem.* **52** 4063
46. Shu-Feng Z, Wang L, Di Y M, Xue C C, Duan W, Li C G and Li Y 2008 *Curr. Med. Chem.* **15** 1981
47. Fardis M, Jin H, Jabri S, Cai R Z, Mish M, Tsiang M and Kim C U 2006 *Bioorg. Med. Chem. Lett.* **16** 4031
48. Artico M, Santo R D, Costi R, Novellino E, Greco G, Massa S, Tramontano E, Marongiu M E, Montis A D and Colla P L 1998 *J. Med. Chem.* **41** 3948
49. Maurin C, Lion C, Bailly F, Touati N, Vezin H, Mbemba G, Mouscadet J F, Debyser Z, Witvrouw M and Cotelle P 2010 *Bioorg. Med. Chem.* **18** 5194
50. Porcari A R, Ptak R G, Borysko K Z, Breitenbach J M, Drach J C and Townsend L B 2000 *J. Med. Chem.* **43** 2457